REMARKS

I. Introduction

In response to the Office Action dated July 27, 2004, claims 1, 3, 8, 9, 11, 16, 17, 19 and 24 have been amended. Claims 1, 3-9, 11-17 and 19-24 remain in the application. Re-examination and re-consideration of the application, as amended, is requested.

II. Claim Amendments

Applicants' attorney has made amendments to the claims as indicated above. These amendments were made solely for the purpose of clarifying the language of the claims, and were not required for patentability or to distinguish the claims over the prior art.

III. Non-Art Rejection

On page (2) of the Office Action, claims 1, 9, and 17 were rejected under 35 U.S.C. §112, second paragraph, as being indefinite, because it was unclear what is a department table.

Applicants' attorney respectfully traverses this rejection. The "department table" is defined in the claims as containing "aggregate information about the retail transactional data," and thus is not unclear. Consequently, Applicants' attorney requests withdrawal of the rejection.

IV. Prior Art Rejections

A. The Office Action Rejections

In paragraphs (1)-(2) of the Office Action, claims 1, 3, 7, 9, 11, 15, 17, 19, and 23 were rejected under 35 U.S.C. §103(a) as being unpatentable over Fayyad et al., U.S. Patent No. 6,263,337 (Fayyad) in view of Eder, U.S. Patent No. 6,321,205 (Eder). In paragraph (3) of the Office Action, claims 4, 6, 8, 12, 14, 16, 20, 22, and 24 were rejected under 35 U.S.C. §103(a) as being unpatentable over Fayyad in view of Eder and further in view of Lazarus et al., U.S. Patent No. 6,430,539 (Lazarus). In paragraph (4) of the Office Action, claims 5, 13, and 21 were rejected under 35 U.S.C. §103(a) as being unpatentable over Fayyad in view of Eder and further in view of Van Huben et al., U.S. Patent No. 6,327,594 (Van Huben).

Applicants' attorney respectfully traverses these rejections.

B. The Applicants' Independent Claims

Independent claim 1 is directed to a data structure for analyzing data in a computer-implemented data mining system, wherein the data structure is a data model that comprises a Gaussian Mixture Model that stores retail transactional data, a basket table that contains summary information about the retail transactional data, an item table that contains information about individual items referenced in the retail transactional data, and a department table that contains aggregate information about the retail transactional data, and the data model is mapped to aggregate the transactional data for cluster analysis of shopping behavior.

Independent claim 9 is directed to a method for analyzing data in a computer-implemented data mining system, comprising: generating a data structure in the computer-implemented data mining system, wherein the data structure is a data model that comprises a Gaussian Mixture Model that stores retail transactional data, a basket table that contains summary information about the retail transactional data, an item table that contains information about individual items referenced in the retail transactional data, and a department table that contains aggregate information about the retail transactional data; and mapping the data model to aggregate the transactional data for cluster analysis of shopping behavior.

Independent claim 17 is directed to an apparatus for analyzing data in a computer-implemented data mining system, comprising: means for generating a data structure in the computer-implemented data mining system, wherein the data structure is a data model that comprises a Gaussian Mixture Model that stores retail transactional data, a basket table that contains summary information about the retail transactional data, an item table that contains information about individual items referenced in the retail transactional data, and a department table that contains aggregate information about the retail transactional data; and means for mapping the data model to aggregate the transactional data for cluster analysis of shopping behavior.

C. The Fayyad Reference

Fayyad describes one exemplary embodiment providing a data mining system for use in finding clusters of data items in a database or any other data storage medium. Before the data evaluation begins a choice is made of the number M of models to be explored, and the number of clusters (K) of clusters within each of the M models. The clusters are used in categorizing the data in the database into K different clusters within each model. An initial set of estimates for a data distribution of each model to be explored is provided. Then a portion of the data in the database is

read from a storage medium and brought into a rapid access memory buffer whose size is determined by the user or operating system depending on available memory resources. Data contained in the data buffer is used to update the original model data distributions in each of the K clusters over all M models. Some of the data belonging to a cluster is summarized or compressed and stored as a reduced form of the data representing sufficient statistics of the data. More data is accessed from the database and the models are updated. An updated set of parameters for the clusters is determined from the summarized data (sufficient statistics) and the newly acquired data. Stopping criteria are evaluated to determine if further data should be accessed from the database.

D. The Eder Reference

Eder describes an automated system and method for evaluating the probable impact of user-specified or system generated changes in business value drivers on the other value drivers, the financial performance and the future value of a commercial enterprise. Value drivers are identified using search algorithms and induction algorithms that define the value drivers associated with each element of the enterprise. After identifying enterprise value drivers the system completes a detailed valuation of the firm using predictive models to determine the relative impact of each value driver on the overall valuation. The detailed valuation results are then used to define a financial simulation model such as a Markov Chain Monte Carlo model. The financial simulation model then analyzes the impact of user specified changes in value drivers on financial performance or generates a list of recommended changes to value drivers that achieve a user specified financial goal.

E. The Lazarus Reference

Lazarus describes predictive modeling of consumer financial behavior by application of consumer transaction data to predictive models associated with merchant segments. Merchant segments are derived from consumer transaction data based on co-occurrences of merchants in sequences of transactions. Merchant vectors representing specific merchants are clustered to form merchant segments in a vector space as a function of the degree to which merchants co-occur more or less frequently than expected. Each merchant segment is trained using consumer transaction data in selected past time periods to predict spending in subsequent time periods for a consumer based on previous spending by the consumer. Consumer profiles describe summary statistics of consumer spending in and across merchant segments. Analysis of consumers associated with a segment

identifies selected consumers according to predicted spending in the segment or other criteria, and the targeting of promotional offers specific to the segment and its merchants.

F. The Van Huben Reference

Van Huben describes a common access method to enable disparate pervasive computing devices to interact with centralized data management systems. A modular, scalable data management system is envisioned to further expand the role of the pervasive devices as direct participants in the data management system. This data management system has a plurality of data managers and is provided with a plurality of data managers in one or more layers of a layered architecture. The system performs with a data manager and with a input from a user or pervasive computing device via an API a plurality of process on data residing in heterogeneous data repositories of computer system including promotion, check-in, check-out, locking, library searching, setting and viewing process results, tracking aggregations, and managing parts, releases and problem fix data under management control of a virtual control repository having one or more physical heterogeneous repositories. The system provides for storing, accessing, tracking data residing in said one or more data repositories managed by the virtual control repository. DMS applications executing directly within, on or behalf of, the pervasive computing device organize data using the PFVL paradigm. Configurable managers include a query control repository for existence of peer managers and provide logic switches to dynamically interact with peers. A control repository layer provides a common process interface across all managers. A command translator performs the appropriate mapping of generic control repository layer calls to the required function for the underlying storage engine.

G. Applicants' Independent Claims Are Patentable Over The References

Applicants' invention, as recited in independent claims 1, 9 and 17, is patentable over the references, because the claims recite limitations not found in the references. Specifically, the combination of Fayyad, Eder, Lazarus and Van Huben does not disclose a data model that comprises a Gaussian Mixture Model that stores retail transactional data, a basket table that contains summary information about the retail transactional data, an item table that contains information about individual items referenced in the retail transactional data, and a department table that contains aggregate information about the retail transactional data, and the data model is mapped to aggregate the transactional data for cluster analysis of shopping behavior.

The Examiner cites Fayyad as teaching most of the elements of the independent claims, including a data structure for analyzing data in a computer-implemented data mining system, as reference number 12 in FIG. 2 and in the accompanying text. The Examiner also cites Fayyad as teaching that the data structure is a data model that comprises a Gaussian Mixture Model that stores transactional data, at col. 9, lines 22-67. In addition, the Examiner cites Fayyad as teaching that the data model is mapped to aggregate the transactional data for cluster analysis, at col. 8, lines 34-46. However, the Examiner admits that Fayyad does not disclose a basket table that contains summary information about the transactional data, an item table that contains information about individual items referenced in the transactional data, and a department table that contains aggregate information about the transactional data. Nonetheless, the Examiner asserts that Eder teaches these elements. Specifically, the Examiner asserts that Eder teaches a basket table that contains summary information about the transactional data at Table 6, col. 13, lines 21-47; an item table that contains information about individual items referenced in the transactional data at Tables 7 and 8, col. 14, line 9 to col. 15, line 13; and a department table that contains aggregate information about the transactional data at Table 10, col. 15, line 35 to col. 16, line 7. Moreover, the Examiner asserts that it would have been obvious to one of ordinary skill to combine Fayyad and Eder to enable the user to group the useful information about transactional data into subgroups and to organize data in the data mining system.

Applicants' attorney disagrees.

First, Applicants' attorney respectfully notes that MPEP §706.02(j) requires that "there must be some suggestion or motivation, either in the references themselves or in the knowledge generally available to one of ordinary skill in the art, to modify the reference or to combine reference teachings." Neither Fayyad and Eder provide such motivation. Indeed, nowhere does the Examiner cite either reference as providing such motivation. Instead, the motivation is provided by the Examiner, which is improper. On this basis alone, the rejection should be withdrawn.

Moreover, at the locations indicated above, Fayyad and Eder, taken individually or in combination, do not teach the claim limitations directed to a data model comprising a Gaussian Mixture Model that stores retail transactional data, a basket table that contains summary information about the retail transactional data, an item table that contains information about individual items referenced in the retail transactional data, and a department table that contains aggregate information about the retail transactional data, and the data model is mapped to aggregate the transactional data for cluster analysis of shopping behavior.

For example, consider the teaching of Fayyad at col. 9, lines 22-67:

> Fayyad: col. 9, lines 22-67
>
> The procedure 120 uses the existing model to create 202 an Old_Model in a data structure like that of FIG. 6D, then determines 204 the length of the pointer arrays of FIGS. 6A-6C and computes 206 means and covariance matrices from the Old_Model SUM, SUMSQ and M data. The set of Old_Model means and covariance matrices are stored as a list of length K where each element of the list includes two parts:
>
> 1) a vector of length n (called the "mean") which stores the mean of the corresponding Gaussian or cluster
>
> 2) a matrix of size n.times.n (called the "CVMatrix") which stores the values of a covariance matrix of the corresponding Gaussian or cluster.
>
> The structure holding the means and covariance matrices are referred to below as "Old_SuffStats".
>
> To compute the matrix CVMatrix for a given cluster from the sufficient statistics SUM, SUMSQ and M (in FIG. 6D), the extended EM procedure computes an outer product defined for 2 vectors OUTERPROD(vector1,vector2). The OUTERPROD operation takes 2 vectors of length n and returns their outer product, or the n.times.n matrix with an entry in row h and column j being vector1(h)*vector2(j). A DETERMINANT function computes the determinant of a matrix. The procedure 200 also uses a function, INVERSE that computes the inverse of a matrix. A further function TRANSPOSE returns the transpose of a vector (i.e. changes a column vector to a row vector). The function EXP(z) computes the exponential e.sup.z.
>
> A function 'ConvertSuffStats' calculates 206 the mean and covariance matrix from the sufficient statistics stored in a cluster model (SUM,SUMSQ,M)
>
> [Mean,CVMatrix]=ConvertSuffStats(SUM,SUMSQ,M)
> Mean=(1/M)*SUM:
> MSq=M*M;
> OutProd=OUTERPROD(SUM,SUM);
> CVMatrix=(1/MSq)*(M*SUMSQ-3*OutProd);
>
> The data structures of FIG. 6A-6D are initialized 100 before entering the FIG. 4 processing loop. In order for the extended EM procedure 120 to process a first set of data read into the memory, the MODEL data structure of FIG. 6D that is copied into Old_Model is not null. An initial set of cluster means is presented to the process. One procedure is to randomly choose the means and place them in the vector 'Sum' and setting M=1.0. For a clustering number K=2 for the data format from Table 1, assume the SUM vector is given as Table 2 for these two clusters.

Nothing in the above discussion of Fayyad teaches "a data structure is a data model that comprises a Gaussian Mixture Model that stores retail transactional data." Indeed, nowhere is retail transactional data found in the above discussion. Instead, the above discussion relates only to a general discussion of computing the mean and covariance matrix of a corresponding Gaussian or cluster. Moreover, while Fayyad generally states that clustering can be used in marketing, his

embodiment refers specifically a database comprised on personal information (age, income, number of children, number of cars owned, etc.

In another example, consider the teaching of Fayyad at col. 8, lines 34-46:

> Fayyad: col. 8, lines 34-46
>
> An additional data structure designated DS in FIG. 6A includes an array of pointers 160 that point to a group of k vectors (the cluster number) of n elements 162 designated 'SUM' a second group of k vectors 164 designated 'SUMSQ', and a group 166 of k floats designated M. This data structure is similar to the data structure of FIG. 6D that describes the MODEL. It contains sufficient statistics for a number of data records that have been compressed into the data structure shown rather than maintained in memory. Compression of the data into this data structure and the CS data structure described below frees up memory for accessing other data from the database at the step 10 on a next subsequent iteration of the FIG. 4 scalable EM process.
>
> A further data structure designated CS in FIG. 6B is an array of c pointers where each pointer points to an element which consists of a vector of n elements (floats) designated 'SUM', a vector of n elements (floats) designated 'SUMSQ', and a scalar 'M'. The data structure CS also represents multiple data points into a vector similar to the MODEL data structure.
>
> A data structure designated RS (FIG. 6C) is a group of r vectors having n dimensions. Each of the vectors has n elements (floats) representing a singleton data point of the type SDATA. As data is read in from the database at the step 110 it is initially stored in the set RS since this data is not associated with any cluster. The current implementation of the extended EM analysis, RS is a vector of pointers to elements of type SDATA of the same length as the 'SUM' vector of the other data structures and a 'SUMSQ' vector is simply null and M=1.

Nothing in the above discussion of Fayyad teaches "the data model is mapped to aggregate the transactional data for cluster analysis." Indeed, nowhere is retail transactional data or the aggregation of retail transactional data or the mapping of a data model to perform the aggregation of retail transactional data found in the above discussion. Instead, the above discussion relates only to two data structures that each contains cluster numbers, SUM elements, SUMSQ elements and M elements; and another data structure that contains a vector of pointers to elements of type SDATA.

In another example, consider the teaching of Eder at Table 6, col. 13, lines 21-47:

> Eder: Table 6, col. 13, lines 21-47
>
> General ledger accounting systems also require that the asset account balances equal the sum of the liability account balances and equity account balances at all times.
>
> The general ledger system generally maintains summary, dollar only transaction histories and balances for all accounts while the associated subsystems, accounts payable, accounts receivable, inventory, invoicing, payroll and purchasing,

maintain more detailed historical transaction data and balances for their respective accounts. It is common practice for each subsystem to maintain the detailed information shown in Table 6 for each transaction.

TABLE 6

Subsystem	Detailed Information
Accounts Payable	Vendor, Item(s), Transaction Date, Amount Owed, Due Date, Account Number
Accounts Receivable	Customer, Transaction Date, Product Sold, Quantity, Price, Amount Due, Terms, Due Date, Account Number
Capital Asset	Asset ID, Asset Type, Date of Purchase, Purchase Price, Useful Life, Depreciation Schedule, Salvage Value
Inventory	Item Number, Transaction Date, Transaction Type, Transaction Qty, Location, Account Number
Invoicing	Customer Name, Transaction Date, Item(s) Sold, Amount Due, Due Date, Account Number
Payroll	Employee Name, Employee Title, Pay Frequency, Pay Rate, Account Number
Purchasing	Vendor, Item(s), Purchase quantity, Purchase Price(s), Due Date, Account Number

Nothing in the above discussion of Eder teaches "a basket table that contains summary information about the retail transactional data." Instead, the above discussion relates only to a general ledger system.

In another example, consider the teaching of Eder at Tables 7 and 8, col. 14, line 9 to col. 15, line 13:

Eder: Tables 7 and 8, col. 14, line 9 to col. 15, line 13

While advanced financial systems are similar between firms, operation management systems vary widely depending on the type of company they are supporting. These systems typically have the ability to not only track historical transactions but to forecast future performance. For manufacturing firms, operation management systems such as Enterprise Requirements Planning Systems (ERP), Material Requirement Planning Systems (MRP), Purchasing Systems, Scheduling Systems and Quality Control Systems are used to monitor, coordinate, track and plan the transformation of materials and labor into products. These systems will generally track information about the performance of the different vendors that supply materials to the firm including the information shown in Table 7.

TABLE 7

Operation Management System Vendor Information

1. Vendor Name
2. Vendor Number
3. Commodity Code(s)
4. Year to date dollar volume
5. Historical dollar volume
6. Percentage of deliveries rejected by QC

7. Percentage of deliveries accepted out of specification
8. Compliance with ISO 9000
9. Actual lead time required for purchases
10. Terms and conditions for purchases
11. Average Delivery Quantity Variance
12. Average Delivery Date Variance
13. EDI* vendor - Yes or No
*EDI = Electronic Data Interchange

Systems similar to the one described above may also be useful for distributors to use in monitoring the flow of products from a manufacturer.

Operation Management Systems in manufacturing firms may also monitor information relating to the production rates and the performance of individual production workers, production lines, work centers, production teams and pieces of production equipment including the information shown in Table 8.

TABLE 8

Operation Management System - Production Information

1. ID number (employee id/machine id)
2. Actual hours - last batch
3. Standard hours - last batch
4. Actual hours - year to date
5. Actual/Standard hours - year to date %
6. Actual setup time - last batch
7. Standard setup time - last batch
8. Actual setup hours - year to date
9. Actual/Standard setup hrs - yr to date %
10. Cumulative training time
11. Job(s) certifications
12. Actual scrap - last batch
13. Scrap allowance - last batch
14. Actual scrap/allowance - year to date
15. Rework time/unit last batch
16. Rework time/unit year to date
17. QC rejection rate - batch
18. QC rejection rate - year to date

Operation management systems are also useful for tracking requests for service to repair equipment in the field or in a centralized repair facility. Such systems generally store information similar to that shown below in Table 9.

Nothing in the above discussion of Eder teaches "an item table that contains information about individual items referenced in the retail transactional data." Applicants' attorney notes that the term "items" is defined in this application as "items purchased by customers." Instead, the above discussion relates only to vendor and production information.

In yet another example, consider the teaching of Eder at Table 10, col. 15, line 35 to col. 16, line 7:

Eder: Table 10, col. 15, line 35 to col. 16, line 7

Sales management systems are similar to operation management systems in that they vary considerably depending on the type of firm they are supporting and they generally have the ability to forecast future events as well as track historical occurrences. In firms that sell customized products, the sales management system is generally integrated with an estimating system that tracks the flow of estimates into quotations, orders and eventually bills of lading and invoices. In other firms that sell more standardized products, sales management systems generally are used to track the sales process from lead generation to lead qualification to sales call to proposal to acceptance (or rejection) and delivery. All sales management systems would be expected to store information similar to that shown below in Table 10.

TABLE 10

Sales Management System - Information

1. Customer/potential customer name
2. Customer number
3. Address
4. Phone number
5. Source of lead
6. Date of first purchase
7. Date of last purchase
8. Last sales call/contact
9. Sales call history
10. Sales contact history
11. Sales history: product/qty/price
12. Quotations: product/qty/price
13. Custom product percentage
14. Payment history
15. Current A/R balance
16. Average days to pay

Nothing in the above discussion of Eder teaches "a department table that contains aggregate information about the retail transactional data." Instead, the above discussion relates only to customer data.

Consequently, the Fayyad and Eder references, taken individually or in combination, do not describe a data model comprising a Gaussian Mixture Model that stores retail transactional data, a basket table that contains summary information about the retail transactional data, an item table that contains information about individual items referenced in the retail transactional data, and a department table that contains aggregate information about the retail transactional data, and the data model is mapped to aggregate the retail transactional data for cluster analysis of shopping behavior.

Moreover, Lazarus and Van Huben fail to overcome these limitations of Fayyad and Eder. Recall that Lazarus was cited only for teaching that the data model is stored in a relational database managed by a relational database management system, and then only against dependent claims 4, 6,

8, 12, 14, 16, 20, 22 and 24, and Van Huben was cited only for teaching a relational database management system for storing the data model, and then only against dependent claims 5, 13 and 21.

Thus, the references do not teach or suggest Applicants' invention. Moreover, the various elements of Applicants' claimed invention together provide operational advantages over the references. In addition, Applicants' invention solves problems not recognized by the references.

Thus, Applicants' attorney submits that independent claims 1, 9 and 17 are allowable over the references.

H. Applicants' Dependent Claims Are Patentable Over The References

Dependent claims 3-8, 11-16 and 19-24 are submitted to be allowable over the references in the same manner as the independent claims, because they are dependent on independent claims 1, 9 and 17, respectively, and thus contain all the limitations of the independent claims. In addition, dependent claims 3-8, 11-16 and 19-24 recite additional novel elements not shown by the references.

With regard to claims 3, 11 and 19, which recite that the cluster analysis groups the retail transactional data into coherent groups according to perceived similarities in the retail transactional data, the Examiner states that Fayyad teaches these limitations at col. 8, lines 35-64. Applicants' attorney disagrees. Instead, Applicants' attorney submits that Fayyad, at the indicated location, merely describes data structures generally, and vectors of elements specifically, but says nothing about grouping retail transactional data into coherent groups according to perceived similarities in the retail transactional data.

With regard to claims 4, 12 and 20, which recite that the data model is stored in a relational database managed by a relational database management system, these claims stand or fall with the independent claims 1, 10 and 18, respectively.

With regard to claims 5, 13 and 21, which recite that the data model is accessed from a relational database managed by a relational database management system, these claims stand or fall with the independent claims 1, 10 and 18, respectively.

With regard to claims 6, 14 and 22, which recite that the data model is mapped into a single flat table format to produce a correct level of aggregation for statistical analysis, the Examiner states that Lazarus teaches these limitations at Table 3 and col. 14, lines 15-51. Applicants' attorney disagrees. Instead, Applicants' attorney submits that Lazarus, at the indicated location, merely

describes a customer transaction file, but says nothing about mapping a data model into a single flat table format to produce a correct level of aggregation for statistical analysis.

With regard to claims 7, 15 and 23, which recite that the data model is mapped into a database view to produce a correct level of aggregation for statistical analysis, the Examiner states that Fayyad teaches these limitations at col. 8, lines 34-44. Applicants' attorney disagrees. Instead, Applicants' attorney submits that Fayyad, at the indicated location, merely describes data structures generally, and vectors of elements specifically, but says nothing about database views or mapping a data model into a database view to produce a correct level of aggregation for statistical analysis.

With regard to claims 8, 16 and 24, which recite that the data model is comprised of one row per transaction in the retail transactional data, the Examiner states that Lazarus teaches these limitations at Table 3 and col. 14, lines 15-51. Applicants' attorney disagrees. Instead, Applicants' attorney submits that Lazarus, at the indicated location, merely a customer transaction file, but says nothing about data models comprised of one row per transaction in the retail transactional data.

V. Conclusion

In view of the above, it is submitted that this application is now in good order for allowance and such allowance is respectfully solicited. Should the Examiner believe minor matters still remain that can be resolved in a telephone interview, the Examiner is urged to call Applicants' undersigned attorney.

Respectfully submitted,

GATES & COOPER LLP
Attorneys for Applicants

Howard Hughes Center
6701 Center Drive West, Suite 1050
Los Angeles, California 90045
(310) 641-8797

Date: October 27, 2004

By: [signature]
Name: George H. Gates
Reg. No.: 33,500

GHG/